

Datasheet for Datasets

Antoine Blanchard

2022-05-29

Motivations pour la création du jeu de données	Composition du jeu de données	Processus de collecte des données	Pré- traitement des données	Diffusion du jeu de données	Maintenance du jeu de données	Considérations légales et éthiques
<ul style="list-style-type: none">• Pourquoi le jeu de données a-t-il été initialement créé ?	<ul style="list-style-type: none">• Que contient le jeu de données principalement ? Les enregistrements représentent-ils principalement des documents, des personnes, des territoires, des entreprises... ?	<ul style="list-style-type: none">• Comment les données ont été collectées (avec des capteurs, manuellement, par des outils informatiques...) ?	<ul style="list-style-type: none">• Comment les données ont-elles été nettoyées ou préparées ?	<ul style="list-style-type: none">• Les données sont-elles diffusées en ligne ? Selon quelles modalités (sur un portail open data, un site web, une API...) ?	<ul style="list-style-type: none">• Qui assure la maintenance du jeu de données ? Comment peut-on contacter cette personne ? Quel est le service responsable du jeu de données ?	<ul style="list-style-type: none">• Si le jeu de données concerne des individus, ont-ils exprimé leur consentement de manière claire ?
<ul style="list-style-type: none">• Quelles ont été les utilisations non prévues du jeu de données ?	<ul style="list-style-type: none">• Dispose-t-on d'un schéma décrivant les variables du jeu de données ?	<ul style="list-style-type: none">• Qui a assuré le processus de collecte de données (des agents, des bénévoles, des étudiants...) ?	<ul style="list-style-type: none">• Les données « brutes » ont-elles été conservées ? Sont-elles diffusées ?	<ul style="list-style-type: none">• Si non, les données sont-elles diffusées au cas par cas ? à la demande ?	<ul style="list-style-type: none">• Est-ce que les rôles sont distincts entre la production des données, leur éditorialisation et leur diffusion ?	<ul style="list-style-type: none">• Les individus ont-ils été informés sur la finalité du traitement de données ?
<ul style="list-style-type: none">• Pour quelles autres tâches le jeu de données pourrait-il être utilisé ?	<ul style="list-style-type: none">• Que contient chaque champ du jeu de données ?	<ul style="list-style-type: none">• Quelle a été la période de collecte des données ?	<ul style="list-style-type: none">• L'outil de pré-traitement des données est-il disponible ?	<ul style="list-style-type: none">• Selon quelle licence les données sont-elles diffusées ?	<ul style="list-style-type: none">• Le jeu de données sera-t-il mis à jour ? Si oui, à quelle fréquence ?	<ul style="list-style-type: none">• Le jeu de données peut-il exposer de manière directe ou indirecte des individus ?

Motivations pour la création du jeu de données	Composition du jeu de données	Processus de collecte des données	Pré- traitement des données	Diffusion du jeu de données	Maintenance du jeu de données	Considérations légales et éthiques
<ul style="list-style-type: none"> • Quelles sont les utilisations trompeuses du jeu de données ? 	<ul style="list-style-type: none"> • Est-ce que le contenu du jeu de données dépend de ressources externes (ex. identifiant SIRET ou lien vers le document...) ? De quelles garanties dispose-t-on concernant la pérennité de ces ressources ? 	<ul style="list-style-type: none"> • Les données ont-elles été collectées directement ou inférées à partir d'autres données ? 		<ul style="list-style-type: none"> • Des redevances ou des restrictions sont-elles appliquées dans l'accès aux données ? 	<ul style="list-style-type: none"> • Si les données deviennent obsolètes, comment cette information sera-t-elle communiquée ? 	<ul style="list-style-type: none"> • Ces données sont-elles conformes au RGPD ?
<ul style="list-style-type: none"> • Qui a financé ou soutenu la création du jeu de données ? 		<ul style="list-style-type: none"> • Les données ont-elles été collectées sur un échantillon ? Quelle est la population complète ? Selon quelles méthodes ? • Quelles sont les erreurs connues, les limites, les sources de bruit ou de redondances associées à ces données ? 			<ul style="list-style-type: none"> • Est-il possible de contribuer à l'amélioration des données ? Selon quelles modalités ? 	<ul style="list-style-type: none"> • Les données peuvent-elles avantager ou désavantager des groupes sociaux ? • Le jeu de données contient-il des informations pouvant être considérées comme inappropriées ou offensantes ?